ST512

Fall Quarter, 2002

Second Midterm Exam

Name: (KEY) ______________________________

Directions: Answer questions as directed. Not that much computation is involved, but show work where appropriate. For true/false questions, circle either true or false.

1. An experiment looks at the amylase specific activity (y) of sprouted maize under 32 treatment conditions that are factorial combinations of three factors:

   - A: analysis temperature (8 levels),
   - B: growth temperature (2 levels)
   - C: variety (2 cultivars).

   There were a total of $N = 96$ experimental units randomized to these 32 treatment combinations. Use $\alpha = 0.05$ in the following questions.

Source	df	SS	MS	F	$p-$value
A	7	327812	46830	72.94	$< .0001$
B		1155	1155	1.80	0.1846
C		63808	63808	99.38	$< .0001$
A*B		7157	1022	1.59	0.1538
A*C		1174	168	0.26	0.9666
B*C		10648	10648	16.58	0.0001
A*B*C		6258	894	1.39	0.2240
Error		41091			
Total		459101			

   (a) True: The third-order interaction ABC is not significant.

   (b) The second-order interaction involving factors A and B is significant. False: *p-value is 0.15*

   (c) There is no evidence of any association between mean amylase specific activity and growth temperature. False: *BC interaction significant, indicating dependence on B*

   (d) True: The effect of factor A, analysis temperature, does not depend significantly on either factor B or factor C.

   (e) True: In the ANOVA table above, $MSE = 41091/64 = 642$.

2. A study was performed to look at systolic blood pressure (BP) in three different diet groups:

   - strict vegetarians (SV)
   - lactovegetarians (LV) (eat dairy products)
   - normal - (standard American diet)

   Samples of size $n = 5$ from each gender were taken from each dietary group and BP was measured. Total sample size is $N = 30$. Let y_{ijk} denote the BP for subject k in diet group i and gender j. Use the SAS output at the end of the problem to answer the questions below:

   (a) Estimate the standard deviation of BPs among people in a given dietary-gender classification group, assuming homoscedasticity across the 6 groups.

   $$\sqrt{MS[E]} = 8.6$$

   (b) True: The $MS[E]$ is the average of the sample variances for the 6 treatment combinations, *or the so-called pooled variance.*

   $$MS[E] = \frac{1}{6}(s_{11}^2 + s_{12}^2 + s_{21}^2 + s_{22}^2 + s_{31}^2 + s_{32}^2)$$

   where $s_{ij}^2 = \frac{1}{4}\sum_{k=1}^{5}(y_{ijk} - \bar{y}_{ij+})^2$.

   (c) Test for a diet×gender interaction effect at level $\alpha = 0.05$.
   From the output, the p-value for interaction is 0.4071 *which is not significant, providing no evidence of interaction.*

   (d) Report a p-value for the null hypothesis that the average BPs in the three diet populations are equal. *From the output, the p-value for DIET is* 0.0223 *which is significant, providing evidence of a diet effect.*

   (e) Estimate the difference between mean BP for men and women.

   $$\bar{y}_{male} - \bar{y}_{female} = 120.1 - 109.2 = 10.9$$

   (f) There is significant ($\alpha = 0.05$) evidence to suggest that population differences among diet groups are different for men than for women: False: *in part c it was determined that there was no evidence of interaction*

   (g) Let $\mu_{NOR}, \mu_{LV}, \mu_{SV}$ denote the mean BPs in the three populations after averaging over gender. Prepare a table of estimates, with estimated standard errors in parentheses:

   $$\begin{aligned} \hat{\mu}_{SV} &= 110.2(2.7) \\ \hat{\mu}_{LV} &= 112.6(2.7) \\ \hat{\mu}_{NOR} &= 121.2(2.7) \end{aligned}$$

   The SE for any DIET mean is $\sqrt{MSE/10} = 2.7$

(h) Suppose you are interested in $c = 4$ contrasts: the 3 pairwise contrasts and the difference θ between mean BP for normal eaters and the average of the mean BPs for the two vegetarian populations. Use Scheffé's procedure to obtain simultaneous 95% confidence intervals for these four contrasts. Note that

$$\begin{aligned}
\sqrt{(3-1)F(0.95,2,24)MS[E]\left(\frac{2}{10}\right)} &= 10.1 \\
\sqrt{(3-1)F(0.95,2,24)MS[E]} &= 22.5 \\
\sqrt{\frac{1}{10}+\frac{1/4}{10}+\frac{1/4}{10}} &= 0.4
\end{aligned}$$

$$\begin{aligned}
\mu_{NOR}-\mu_{SV}: &\quad 11.0 \pm 10.1 \\
\mu_{NOR}-\mu_{LV}: &\quad 8.6 \pm 10.1 \\
\mu_{LV}-\mu_{SV}: &\quad 2.4 \pm 10.1 \\
\theta: &\quad 9.8 \pm 9
\end{aligned}$$

For the last contrast,

$$SE = \sqrt{MS[E]\left(\frac{1}{10}+\frac{1/2^2}{10}+\frac{1/2^2}{10}\right)} = 22.5 \times 0.4 = 9.0$$

(i) Label each sum of squares below so that the source of variability which it quantifies is clear

$$SS[diet], SS[gender], SS[diet \times gender], SS[E], SS[Tot]$$

Sum of squares	Label
$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{5}(y_{ijk}-\bar{y}_{ij+})^2$	$SS[E]$
$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{5}(\bar{y}_{i++}-\bar{y}_{+++})^2$	$SS[diet]$
$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{5}(\bar{y}_{ij+}-\bar{y}_{i++}-\bar{y}_{+j+}+\bar{y}_{+++})^2$	$SS[diet \times gender]$
$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{5}(\bar{y}_{+j+}-\bar{y}_{+++})^2$	$SS[gender]$
$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{5}(y_{ijk}-\bar{y}_{+++})^2$	$SS[Tot]$

```
proc glm;
   class DIET GENDER;
   model bp=DIET|GENDER;
   means DIET|GENDER;
run;
```


Class Level Information

Class	Levels	Values
DIET	3	LV NOR SV
GENDER	2	female male

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1705.066667	341.013333	4.56	0.0046
Error	24	1793.600000	74.733333		
Corrected Total	29	3498.666667			

R-Square	Coeff Var	Root MSE	BP Mean
0.487348	7.539108	8.644844	114.6667

Source	DF	Type I SS	Mean Square	F Value	Pr > F
DIET	2	669.0666667	334.5333333	4.48	0.0223
GENDER	1	896.5333333	896.5333333	12.00	0.0020
DIET*GENDER	2	139.4666667	69.7333333	0.93	0.4071

The GLM Procedure

Level of DIET	N	BP Mean	BP Std Dev
LV	10	112.600000	7.3665913
NOR	10	121.200000	10.3365586
SV	10	110.200000	12.3809890

Level of GENDER	N	BP Mean	BP Std Dev
female	15	109.200000	9.87927123
male	15	120.133333	9.39503415

Level of DIET	Level of GENDER	N	BP Mean	BP Std Dev
LV	female	5	110.000000	7.8740079
LV	male	5	115.200000	6.5726707
NOR	female	5	115.200000	8.3186537
NOR	male	5	127.200000	9.0111043
SV	female	5	102.400000	10.3344085
SV	male	5	118.000000	9.2736185

3. A common method of assessing cardiovascular capacity is through treadmill exercise testing. Maximal oxygen uptake, or VO_2 max, is an index measured using a treadmill with an inclined protocol. A random sample of $n = 12$ adults was randomized to two treatment groups:

- T_1 : 12-week step-aerobic training program
- T_2 : 12-week outdoor running regimen on flat-terrain.

VO_2 max is measured before and after treatment and the change y is used as the response. Ages of subjects are also measured. Data appear below. The point of the study is to see if the first treatment group shows greater increases in VO_2 max than the 2nd.

Treatment	Age	y	Treatment	Age	y
1	31	17.05	2	23	-0.87
1	23	4.96	2	22	-10.74
1	27	10.4	2	22	-3.27
1	28	11.05	2	25	-1.97
1	22	0.26	2	27	7.5
1	24	2.51	2	20	-7.25
avg	25.8	7.7	avg	23.2	-2.8

An ANCOVA model was fit using SAS (output at end of problem.)

(a) Use an independent variable z_j denoting age of subject j and an indicator variable x_j for step-aerobic group to propose a multiple linear regression model for mean change in VO_2 max for subject j. (Don't fit any models here, just propose one using regression coefficients.)

$$\mu_j = \beta_0 + \beta_1 x_j + \beta z_j \text{ for } j = 1, 2, \ldots, 12.$$

$$x_j = \begin{cases} 0 & \text{flat-terrain group} \\ 1 & \text{step-aerobics group} \end{cases} \quad \text{for } j = 1, 2, \ldots, 12.$$

(b) Use the SAS output to report the estimated mean change in $V0_2$ max separately for each group as a function of age. (Estimate the model you specified in part a)

$$\begin{aligned} \hat{\mu}(x, z) &= \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta} z \\ &= \begin{cases} -41.02 + 1.89z & \text{step-aerobics} \\ -46.46 + 1.89z & \text{flat-terrain} \end{cases} \end{aligned}$$

(c) Use the model to complete the table of adjusted and unadjusted means. Use the fact that the average age among all subjects is $\bar{z} = 24.5$. Refer to the original dataset on the previous page.

Trt. group	Unadjusted mean	Adjusted mean	
Step-aerobic	7.7	$-46.46 + 5.44 + 1.89(24.5)$	$= 5.29$
Flat-terrain	-2.8	$-46.46 + 0 + 1.89(24.5) =$	-0.16

(d) Is there evidence that mean change in VO_2 max depends on age? Explain briefly.

$$t_\beta = \frac{\hat{\beta} - 0}{SE(\hat{\beta})} = 6.4(p - \text{value} = 0.0001)$$

(so that $\hat{\beta}$ differs significantly from 0, and there is evidence of linear association between age and mean change in VO_2 max)

(e) Estimate the treatment effect and its standard error.

$$\hat{\beta}_1 = 5.44(SE = 1.80)$$

(f) Draw a conclusion regarding change in VO_2 max brought about by the two exercise regimens. *Step-aerobics bring about significantly greater increases in VO_2 max. (Flat-terrain may not bring about any increase.) Further, age helps in explaining variation in change in VO_2 max. On average, older folks are able to benefit more from the exercise regimens than younger folks, at least over the age ranges considered here.*

```
data one;
   input grp z y @@;
   if grp=1 then x=1;
   if grp=2 then x=0;
   cards;
   1 31 17.05 1 23 4.96 1 27 10.4 1 28 11.05 1 22 0.26 1 24 2.51
   2 23 -0.87 2 22 -10.74 2 22 -3.27 2 25 -1.97 2 27 7.5 2 20 -7.25
;
run;
proc reg;
   model y=x z;
run;
```

```
                              The SAS System                                 1
                             The REG Procedure
                           Analysis of Variance
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	647.87492	323.93746	41.42	<.0001
Error	9	70.38877	7.82097		
Corrected Total	11	718.26369			

Root MSE	2.79660	R-Square	0.9020
Dependent Mean	2.46917	Adj R-Sq	0.8802
Coeff Var	113.26091		

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-46.45650	6.93653	-6.70	<.0001
x	1	5.44262	1.79645	3.03	0.0143
z	1	1.88589	0.29534	6.39	0.0001